Selected Topics on Time-Frequency Matrix Decomposition Analysis

Behnaz Ghoraani

Biomedical Engineering, Rochester Institute of Technology, NY, USA

bghoraani@ieee.org

***Abstract*-While time-frequency (TF) representations are commonly used for visualization purposes, the adaptive feature extraction from TF matrices has not been extensively studied in the literature. Even when used, there exists no unique or automated methodology to extract discriminative TF features from non-stationary signals. The present paper focuses on feature extraction from time-frequency distribution (TFD), and attempts to develop a generalized TF matrix (TFM) analysis methodology that exploits the benefits of TFD in pattern classification systems. The proposed TFM feature extraction consists of two stages: first, we build TFM decomposition that uses a matrix decomposition (MD) technique to effectively segment non-stationary signals. Second, instantaneous and unique features are extracted from each segment in a way that they successfully represent joint TF structure of the signal. In this paper, the suitable tools for TFM feature extraction pertaining to pattern classification are investigated through experiments with different synthetic signals. Experiments performed with pathological speech and environmental audio signals produced 98.6% and 85.5% accuracy rates respectively demonstrating the benefits of TFM feature extraction for pattern classification related applications**.

Keywords- Time-Frequency Matrix Analysis; Time-Frequency Analysis; Time-Frequency Feature Extraction; Matrix Decomposition

I. INTRODUCTION

Time-frequency representations are potentially successful features for accurate classification of practical signals. Due to the non-stationary nature of practical signals, the spectral contents of these signals are variable over time. The time domain cannot reveal any explicit information about this non-stationarity. Fourier representation reveals spectral constitutive features of the signal without preserving any explicit localization in time. Hence neither time-domain nor frequency-domain analysis might be sufficient to analyze signals with time-varying frequency contents.

Joint TF distribution (TFD) indicates a two-dimensional energy representation of a signal in terms of time and frequency domain. This characteristic of joint TF domain makes this approach suitable for revealing the non-stationary behavior of practical signals such as trends, discontinuities, and repeated patterns where time or frequency signal processing approaches fail or are not effective [1]. Motivated by the representative characteristics of the TF distributions for real-world signals, the present paper attempts to develop a generalized approach for the application of TF analysis in pattern recognition applications.

The general methodology of a classification involves extracting representative features from a signal and feeding them into a pattern classifier. The better and more effectively features are extracted, the higher performance will be achieved in the classification stage. However, due to the dimensionality of TF representation, the direct application of TFD as features is computationally expensive and not always efficient. For example, for a 64 ms signal with a sampling frequency of 16 kHz, a TFD with 512×1024 resolution contains 524,288 TF samples, which encourages a significant reduction in the dimensionality of the TFD. In addition to the computational complexity issue, employing the entire TF data in the classification stage will customize the designed features to that specific data, and the system might loss its generality to new signals. Therefore, in order to make TFD suitable as TF features, one possible approach is to reduce the dimensionality of the TF distribution by removing the redundant information in the TF plane and preserving its representative region [2, 3, 4]. The dimensionality reduction of TFD to obtain appropriate TF features, is termed as *TF Feature Extraction*. It is worth to mention that there are some approaches that attempt to optimize the TFD and consider all the time-frequency information as features with no dimension reduction [5, 6]. However, our main objective in this manuscript focuses on TF dimension reduction to obtain meaningful features with high temporal/spectral localization that will help in performing signal classification tasks.

Before explaining our proposed TF analysis technique in this paper, we elaborate on the properties of a desirable TF feature extraction as follows:

(i) *Non-stationary compatible:* To characterize the time-varying spectral properties of a signal without any assumptions regarding the stationarity over time. As mentioned before, the reason for utilizing TFD in classification of practical signals is taking advantage of TF plane in revealing the dynamic variations of practical signals. Therefore, it is necessary for any proposed TF feature extraction method to extract features that reflect the non-stationary behavior of real-world signals. Failure in this mission will result in features that do not represent the true characteristics of signals and will subsequently increase the classification error. This is a serious problem particularly in cases of real-world applications where signals contain abrupt changes or discontinuities.

(ii) Long-term analysis: To accurately quantify the information hidden in long durations of signals. A suitable TF analysis should be capable of analyzing a long duration of the signal to select the abnormality regions rather than look for the abnormality

structure over short segments. Such a long-term TF analysis captures the long-term information in a data and improves the recognition and decision making performances. For example, in biomedical signal classification, the abnormality structure may be concentrated in a short duration of the signal. As a result, not all the features obtained from the short segments may represent the signals' abnormality behavior. However, in the classification stage, all the features, which are extracted from the abnormal signals, are assumed to provide a clear discrimination between normal and abnormal signals. Therefore, the classification using short-term features may result in poor classification accuracy. On the other hand, a long-term TF analysis, if is appropriately applied to the signal, has the potential to derive the real abnormality features as related to the abnormality structure.

Another advantage of long-term analysis is providing an understanding about the connectivity of subsequent segments, which may improve the classification performance. In manual data screening and decision-making, the human perception requires at least a few seconds to process a given data to better understand the underlying information. For example, an average individual has to listen to at least 0.5 s to 1 s of an audio to understand the content, and even to learn more about the details in the audio, the person requires listening to a longer episode of that signal. Similarly in automated signal classification, when a longer signal is analyzed, a better understanding of the signal characteristics will be captured in the extracted features, which may enhance the classification accuracy.

(iii) Localized in time and frequency: To provide highly representative TF features. In order to enhance the classification accuracy, the TF features should preserve the localization of the TFD in both frequency and time domain, passing the representative features to the classification stage.

(iv) Dimension reduction: To effectively reduce the dimensionality of the TF plane. In order to reduce the computational complexity, an appropriate TF feature extraction technique should represent the TF structure in a low-dimensional feature space.

The flowchart in Fig. 1 displays the evolution of such a TF feature extraction technique. The structure and contributions of the present paper are as follows: In Section II, we explain a background of TF analysis. We put the TF feature extraction into a new perspective by categorizing this domain into three different TF approaches: TF Approach 1, TF Approach 2, and TF Approach 3. We investigate the pros and cons of each TF approach as related to the TF features' properties of interest explained in Section I. The selected TF approach which we call TF matrix (TFM) feature extraction is described in Section III. TFM feature extraction consists of two stages: TFM decomposition and TF feature extraction. Selection of the right matrix decomposition (MD) for TFM feature extraction and classification is explained in Section IV. Three synthetic experiments are performed to evaluate the performance of three well-known MD techniques in the following stages: TFM decomposition, TFM feature extraction, and the classification accuracy. According to the authors' knowledge, the present paper is the first work that investigates selection of an appropriate MD technique for TF feature analysis. Calibration of the selected MD technique for TFM feature extraction is explained in Section V, where we introduce a novel seeding technique for the MD technique based on the information in the TF plane. Section VI provides the application of the proposed TFM feature extraction methodology in two real world datasets, and the paper is concluded in Section VII.

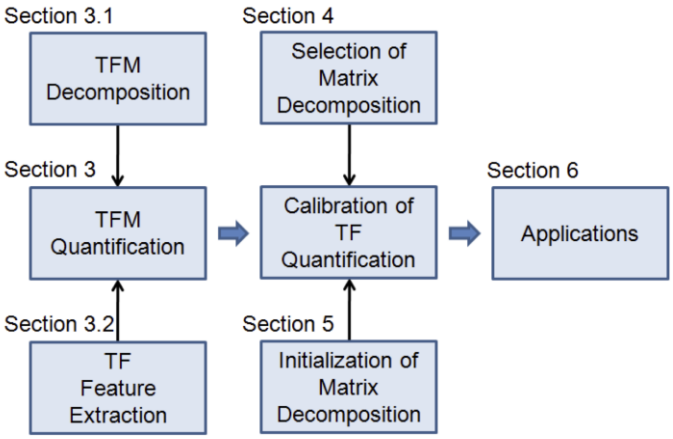


Fig. 1 The flowchart of the proposed TFM feature extraction technique

II. TF FEATURE EXTRACTION BACKGROUND

There are two important directions in the development of TF analysis techniques: *'visualization of the event of interest'* and *'feature extraction and classification of the event of interest'*. The former direction focuses on the TF analysis for representation of an event of interest, and has been widely used in biomedicine including: representation of event-related potential (ERP) activities [7], and visualization of the temporal dynamics of the brain activities and interactions [8, 9, 10]. The latter research direction focuses on adaptation of TF feature extraction for pattern recognition, which is the main focus of the present paper.

TF feature extraction for signal classification and decision-making attempts to reduce the dimensionality of the feature space by removing the redundancy and keeping only the representative contents of the TFD. For a comprehensive review of

these approaches the following references are recommended: [11, 12]. In the present section, we briefly review three available TF feature extraction approaches as related to our contribution. Based on this review, we select an appropriate TFD feature extraction technique as related to the properties of interest listed in the Introduction.

As mentioned above, several TF feature extraction methods have been proposed in the literature. These methods can be grouped into three different approaches as follows:

Approach 1 extracts TF features through statistical or averaging techniques [3, 13]. First, a signal is divided into short windows. Next, a few features are extracted from each window. Although this approach reduces TF dimensionality, the nature of the TF features may not represent signal variations over the analysis window. The feature extraction procedure contains a form of averaging, which lacks instantaneous and non-stationarity information in the signal.

Approach 2 can be applied on relatively long durations of a signal, as the method does not require any stationarity assumption about the signal. This approach interprets the TFD as a matrix (**V**, called TF matrix (TFM)), and employs a matrix decomposition (MD) technique to decompose the TFD into its basis components [14]. These components are used as TF features because they represent spectral variations of the signal over the analysis window. The advantage of this approach is that the TF features are adaptive to non-stationary spectral behavior of the signal. The drawback of this approach is that the length of each feature vector is proportional to the TFD dimension, and therefore, the dimensionality of the TF features may remain very large.

Approach 3 is illustrated in Fig. 2. This method is a combination of the first two approaches. In the first block, TF Approach 2 is used to decompose the TFM into its spectral and temporal vectors; and TF Approach 1 is employed in the second block to further decrease the dimensionality by extracting some statistical features from each decomposed component. In addition to a significant dimension reduction, the extracted TF features provide an efficient representation of signal non-stationarity. This is because of the fact that the approach does not take any assumption about the stationarity of the signal, and MD automatically decides at which intervals the signal is stationarity. Considering the potential benefits of TF Approach 3 for classification applications, a more thorough investigation of this approach is required. Few examples of applications of this approach include: the work of Groutage and Bennink [15, 16], in which the authors apply singular value decomposition (SVD) to the TFD constructed from acoustic signatures for the underwater vehicles. The authors showed that the extracted features localize the stair-step pattern of energy density of the transient signal, but, they did not actually use the TFM features for classification purposes.

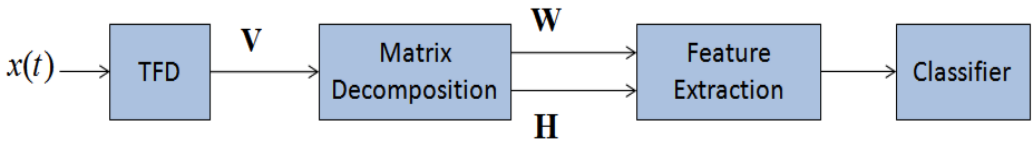


Fig. 2 The general block diagram of the third TF approach, which is called TF matrix (TFM) feature extraction. The matrix decomposition stage makes the approach to be long-term and non-stationary adaptive while reducing the redundant information in the TF plane. The feature extraction stage further reduces the dimensionality of the TFD

Table 1 summarizes the properties of the three TF approaches as related to the properties of a suitable TF feature extraction as mentioned in Section I. Considering the advantages of TF Approach 3 to the other approaches, and as our goal to develop a generalized and effective TF feature extraction approach, in the present study, we expand this approach to extract representative and localized TF features. This TF approach has no theoretical limitations for the utilized TFD. However, utilizing a high resolution and cross-term free TF representation will improve the TF features as better representatives of a signal: features with better localization in time and frequency; and good non-stationary representation.

TABLE 1 DESIRABLE PROPERTIES OF EACH TF FEATURE EXTRACTION APPROACH EXPLAINED IN SECTION II. THE MORE ARE THE NUMBER OF THE ASTERISK AT EACH PROPERTY INDICATES THAT THE METHOD IS MORE DESIRABLE WITH RESPECT TO THAT SPECIFIC PROPERTY

Property	Long-term analysis	Non-stationary compatible	Localized in time and frequency	Dimension reduction
TF Approach 1	*	*	*	***
TF Approach 2	***	**	**	*
TF Approach 3	***	***	***	**

III. TFM FEATURE EXTRACTION

In TFM feature extraction methodology, we treat the TF distribution (V *(f, t)*) as a matrix, and denote the obtained TF matrix (TFM) with **V**M×N, where *N* is the sample length and *M* is the frequency resolution. As shown in Fig. 2, the TFM feature extraction consists of two stages: TFM decomposition and TF feature extraction, which are explained in this section.

A. TFM Decomposition

In this stage, matrix decomposition (MD) technique is applied to the TFM in such a way that TF matrix (**V**M×N) is decomposed as shown below:

$$V_{M\times N} = W_{M\times r} H_{r\times N}$$

$$= \sum_{j\text{-}1}^{r} w_{M\times 1}^{j} h_{1\times N}^{j} \tag{1}$$

In Eq. 1, the MD technique reduces the TFM (**V**) to the base and coefficient vectors ($\{w_{Mx1}^{j}\}_{j=1,\ldots,r}$ and $\{h_{1xN}^{j}\}_{j=1,\ldots,r}$, respectively.) in a way that the former represents the significant frequency structures in the TFD, and the latter specifies the location of each component in time plane. The arrangement of each matrix is shown in the following equation:

$$W_{M\times r} = [w^1\ w^2\ \ldots\ w^r], \tag{2}$$

$$H_{r\times N} = \begin{bmatrix} h^1 \\ h^2 \\ \vdots \\ h^r \end{bmatrix}, \tag{3}$$

We call matrices *W* and *H*, base and coefficient matrices, respectively. The decomposition order, *r*, is a finite number, chosen to be smaller than *M* or *N* so that *W* and *H* are smaller than the original matrix *V*. The efficiency of TFM feature extraction heavily depends on the quality of the decomposed matrices in the TFM decomposition stage, and the quality of the matrix decomposition (MD) technique. Therefore, in Section IV, we investigate through three well-known MD techniques, and select the most appropriate MD technique for the TFM feature extraction.

B. TF Feature Extraction

The next stage in Fig. 2 is extracting TF features from the decomposed base and coefficient vectors. Joint TF moments of a TFD carry important information about the TF characteristics of the signal and have been used for classification of time-varying signals [3] and feature identification [15]. For signal *x (t)*, the joint TF moments are given by

$$\left\langle n^p m^q \right\rangle = \sum_{n=0}^{N}\sum_{m=0}^{M} n^p m^q \mathrm{V}(m,n) \tag{4}$$

where, $V_{\mathrm{M\times N}}$ is the TFM of the signal, and *p* and *q* are the time and frequency moment orders. The joint TF moments can be used to yield the characteristic function, or moment-generating function, *M (θ, τ)*, which can be expanded in a Taylor series given as

$$M(\theta,\tau) = \sum_{p=0}^{\infty}\sum_{m=0}^{\infty} \frac{(j\tau)^p (j\theta)^q}{p!q!} \left\langle n^p m^q \right\rangle \tag{5}$$

Keeping all the joint moments preserves all of the information in the TFD; however, depending on the application, one should decide on how many of the joint moments are required in order to preserve the major portion of the TF information.

Similarly, in the present work, we extract the temporal and spectral moments from the decomposed bases and coefficients as representative features. If we substitute the TFM, *V (m, n)*, in Eq. 4 with its equivalent in Eq. 1, the following formula is obtained:

$$\left\langle n^p m^q \right\rangle = \sum_{n=0}^{N}\ \sum_{m=0}^{M} n^p m^q \sum_{i=1}^{r} w_i(m) h_i(n) \tag{6}$$

Since n^{p} and m^{q} are always non-negative terms, if $w_{\mathrm{i}}(m)\ h_{\mathrm{i}}(n)$ is positive for $i = 1, \ldots, r$ then Eq. 6 can be re-written as

$$\left\langle n^p m^q \right\rangle = \sum_{i-1}^{r}\sum_{n-0}^{N} n^p h_i(n) \sum_{m-0}^{M}\ m^q w_i(m)$$

$$= \sum_{i-1}^{r} \left\langle n_i^p \right\rangle \left\langle m_i^q \right\rangle \tag{7}$$

Where $< m_i^q >$ and $< n_i^p >$ are the moments of the base and coefficient *i* respectively. From Eqs. 5 and 7, it can be seen that if the base and coefficient vectors are non-negative, the joint TF moments can be determined from the decomposed vectors as derived in Eq. 7. The non-negativity of the decomposed vectors is required because *r* is a finite number. If *r* is infinite, the terms in Eq. 6 can be reordered without any condition on the sign. However in a MD application, *r* is selected to be smaller than *M* or *N* so that *W* and *H* are smaller than *V* while containing the most dominant components in the original matrix.

The joint TF moments obtained from decomposed vectors (Eq. 7) are better descriptor of the signal compared to the TF moments extracted from the original matrix (Eq. 4). This is because the former is related to the most dominant characteristics of the signal, while the latter represents the entire TFD including both the representative structure and the redundant information in the TFD. Therefore, instead of calculating the joint TF moments, we determine the moments of the base and coefficient vectors separately, and use them as the TF features of the signal.

IV. MATRIX DECOMPOSITION (MD) TECHNIQUES

In this section, we investigate well-known MD techniques as related to TFM feature extraction, and select the most suitable technique to be used in the proposed pattern recognition system. There are several MD techniques available including principal component analysis (PCA), independent component analysis (ICA), and non-negative matrix factorization (NMF). The objective of PCA is to find a set of orthogonal components that minimize the mean square error of the reconstructed data, and represent the original data with fewer components to reduce the dimensionality of the data [17]. ICA is a statistical technique for decomposing a dataset into components that are as independent as possible [18]. NMF decomposes a non-negative matrix, and constraints the matrix factors *W* and *H* to be non-negative [19].

Performance comparison of ICA and PCA to NMF for feature extraction has been experimentally investigated for different applications; however, depending on the application and data type, different results have been reported. Some works demonstrated the advantages of ICA to NMF and PCA [20, 21, 22], while a conflicting result is reported in some other papers [23, 24]. Considering this contradictory report on the performances of the different MD techniques and lack of a comprehensive comparison among them as related to the TF feature extraction, this section investigates which MD is more appropriate for the TFM feature extraction. Next section performs a fair comparison between the above three well-known techniques on a controllable dataset to select the most suitable technique for feature extraction of TF plane.

Various strategies have been proposed for PCA, ICA, and NMF. In this study, we use Robust PCA (RPCA) which estimates the eigenvalues and eigenvectors using projection pursuit [25], and is more robust than outliers. Fast ICA (Fast ICA) [18] is used for implementation of ICA technique to prevent the algorithm from converging to the same component after iteration. A projected gradient bound-constrained optimization method proposed by Lin [26] is used for NMF optimization. The gradient-based NMF is computationally competitive and offers better convergence properties than the other approaches.

A. Selection of Matrix Decomposition Technique

A successful TFM feature classification method depends on the selection of an appropriate MD technique. In this section, we evaluate Fast ICA, Robust PCA, and NMF for the application of TFM feature extraction and classification. The evaluation is performed at three stages in Fig. 2: *TFM decomposition*, *TFM feature extraction*, and *TFM feature classification*.

TFM decomposition: in Experiment 1, we explore which MD method provides the most accurate decomposition at TFM decomposing stage. *TFM feature extraction*: in Experiment 2, the TF features, which are obtained using different MD methods, are compared as related to their TF localization performances. *TFM feature classification*: Experiment 3 compares the TFM feature classification performance using a well-known synthetic classification problem. Finally, based on the results obtained in these three experiments, the suitable MD technique is selected.

1) Experiment 1: TFM Decomposition

The three MD techniques are used to decompose the TFM of a synthetic dataset. First, a set of synthetic signal is generated using the following equation:

$$x(t) = \sum_{j=1}^{7} x_j(t) = \sum_{j=1}^{7} \alpha_j g(\sigma_j, \mu_j) \sin(2\pi f_j(t)\, f_s t) \tag{8}$$

where for each component (j), α_j is the amplitude parameter, σ_j and μ_j define the duration and location in time, $f_j\,(t)$ is the normalized instantaneous frequency, and f_s is the sampling frequency. In this example, anon-stationary signal is composed of seven Gaussian functions modulated with sinusoidal signals. The parameters ($\alpha, \sigma, \mu, a, b$) are selected so that the synthetic signal contains three common non-stationary structures existing in real-world signals: frequency localized, transient, and frequency modulated components. In case of a frequency localized and transient component, the normalized instantaneous frequency ($f_j\,(t)$ in Eq. 8) is a constant value with respect to time ($\alpha_j = 0$ and $f_j\,(t) = b_j$). For a frequency modulated component, $f_j\,(t)$ is variable over time: $f_j\,(t) = a_j t + b_j$. Table 2 lists the values of the parameters in Eq. 8.

Fig. 3(a) displays a synthetic signal generated as explained above. Fig. 3(b) shows the corresponding TFM of this signal. A hundred synthetic signals are generated. RPCA, FICA and Gradient-based NMF are used to decompose the TF matrix to its **W** and **H** components. The extracted base and coefficient matrices are used to reconstruct the TFM, denoted by **V**rec = **WH**. Using the reconstructed matrix and the pre-known properties of each synthetic signal, we measure the decomposition performance of each MD technique as explained below:

TABLE 2 SELECTION OF THE PARAMETERS IN THE SYNTHETIC DATASET GENERATED USING EQ. 8. U (P, Q) IS A UNIFORM DISTRIBUTION OVER THE RANGE OF [P, Q]

Component (j)	α	σ	μ	a	b
1	1	28	U(1,1024)	0	U(0,0.5)
2	2	2	U(1,1024)	0	U(0,0.5)
3,4	1	32	U(1,1024)	0	U(0,0.5)
5,6	1	2	U(1,1024)	0	U(0,0.5)
7	1	72	U(1,1024)	U(-5,5)	U(0,0.5)

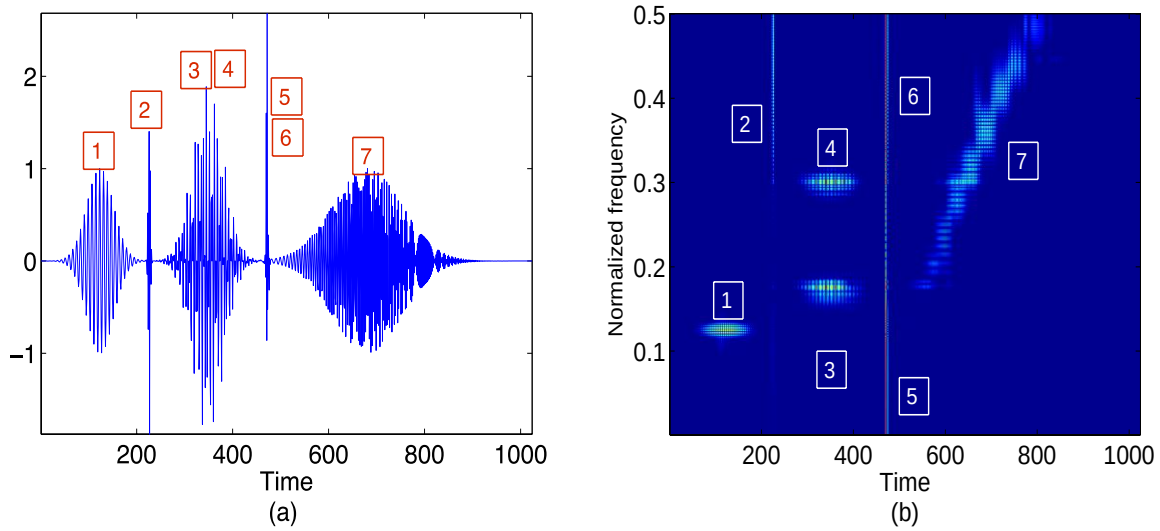


Fig. 3 A synthetic signal in Experiment 1: Eq. 8 is used to generate a synthetic signal. (α, σ, μ, a, b) for each component from 1 to 7 are as following: (1,28,124, 0, 0.125), (2,2,224, 0, 0.375), (1,32,348,0, 0.1750), (1,32,348, 0, 0.3), (1,2,472,0, 0.075), (1,2,472,0, 0.425), (1,72,680,-5, 0.5). The sampling frequency is 4 kHz. (a) The synthetic signal in time domain. (b) TFM of the synthetic signal. Matching pursuit TFD [27] is used to construct the TFM

A moment-based figure of merit is proposed. This measure intends to calculate the reconstruction accuracy according to both the structure and energy of the samples. For each component, we define the first order moment of the reconstructed TF matrix around its normalized instantaneous frequency (f_j (t)) as the localization measure of that component as shown here:

$$Lcz_j = \sum_{n=t_{j1}}^{t_{j2}} \sum_{m=f_j(n)+\Delta}^{f_j(n)-\Delta} \left(\left| V_{rec}(m,n) \right| \times \left| m - f_{j(n)} \right| \right) \tag{9}$$

where 2Δ is the frequency interval around instantaneous frequency $\left(f_j(t)\right)$. t_{j1} *and* t_{j2} are the beginning and ending time of component j. The beginning time of each component is defined as the sample where the amplitude arrives to the 0.1% of the component's maximum amplitude. The same rule applies for the ending time. The percentage localization of component j is calculated as follows:

$$\text{Localization}_j(\%) = 100 - \left(\frac{\left| Lcz_j - Lcz_{O-j} \right|}{Lcz_{O-j}} \times 100 \right) \tag{10}$$

where, Lcz_j and Lcz_{O_j} are the localization values calculated using Eq. 9 from the reconstructed TFM ($\mathbf{V}_{rec}$) and the original TFM ($\mathbf{V}$), respectively. Further detail on this moment-based figure of merit is explained in Appendix A.

Finally, the average localization percentage is computed for the generated synthetic signals, and the results are plotted in Fig. 4. The first three bars show the average localization percentage for the frequency-localized components (components 1, 3 and 4 in Fig. 3 (b)), respectively. The next three bars represent components number 2, 5 and 6, and the last three bars belong to component 7.

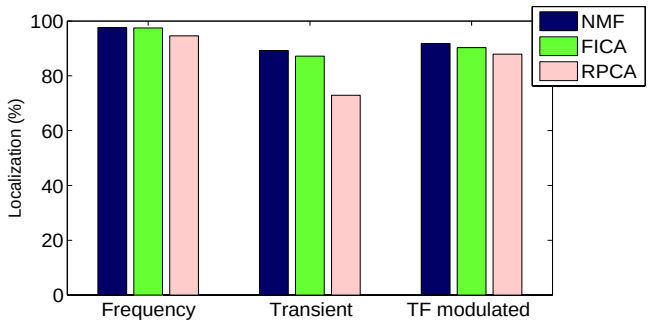


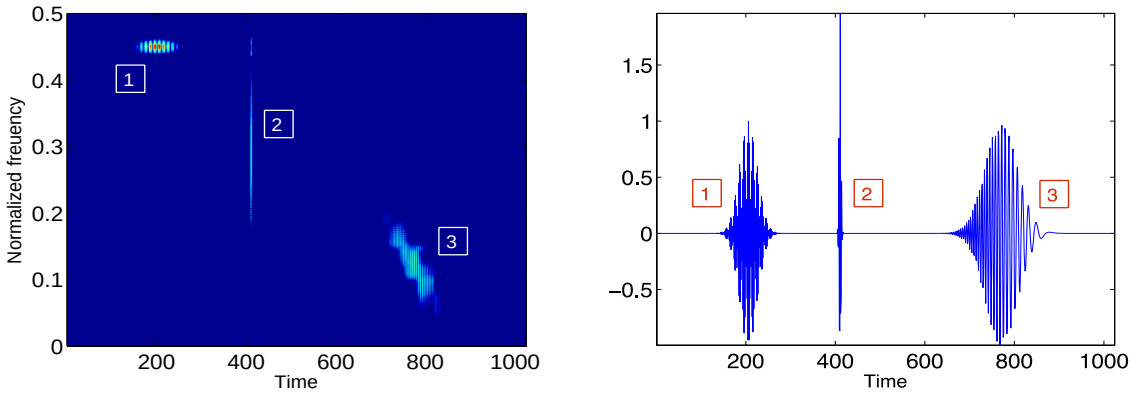
Fig. 4 Result of Experiment 1: Localization performance of NMF, ICA and PCA are compared for frequency localized, transient and frequency modulated components

2) Experiment 2: TF Feature

This experiment compares the performance of the three MD techniques at TF feature extraction stage. The synthetic signal shown in Fig. 5(a) is used in this experiment. This signal is constructed of three components: a frequency localized component, a transient and a frequency linearly modulated component. The TFM features are extracted as explained in the following steps: First, the TFM, $\mathbf{V}_{M\times N}$, is constructed as illustrated in Fig. 5(b). Second, TFM decomposition is performed using RPCA, FICA and gradient-based NMF. Next, TF features are extracted from the decomposed matrices (**W** and **H**). As explained in Section B., spectral and temporal moments should be extracted from non-negative vectors. Since PCA and ICA techniques do not guarantee the non-negativity of decomposed matrices instead of directly using them, their squared values ($\widetilde{\mathbf{W}} = \mathbf{W}.\mathbf{W}^{\mathbf{T}}$ and $\widetilde{\mathbf{H}} = \mathbf{W}.\mathbf{W}^{\mathbf{T}}$, respectively) are used in Eq. 7 [15]. The TF features are extracted as shown in the following:

$$\begin{aligned} F_j &= \left(\hat{e}_j, \bar{n}_j, \bar{m}_j, \hat{n}_j, \hat{m}_j\right) \\ &= \left(10\log(e_j), \langle n\rangle_j, \langle m\rangle_j, \sqrt{\langle n^2\rangle_j - \bar{n}^2{}_j}, \sqrt{\langle m^2\rangle_j - \bar{m}^2{}_j}\right), \end{aligned} \tag{11}$$

where, $< m^p >_j$ and $< n^p >_j$ are the p order spectral and temporal moments, respectively, and e_j is the average energy of component j normalized respect to the component with the maximum energy.



(a) Synthetic signal in time domain (b) TFD of the synthetic signal

Fig. 5 Synthetic signal in Experiment 2: (a) A synthetic signal with three components: from left to right the components are frequency localized, transient and frequency modulated. (b) TFM of the synthetic signal

Finally, Tables 3, 4 and 5 present the derived features utilizing RPCA, ICA and NMF respectively. The feature vectors, F1 to F7, are sorted from the highest energy to the lowest energy. The number of features is chosen based on the energy of each feature. A feature will be considered if its energy is greater than -10dB [15]. For NMF features, the energy after the 7th feature is very small (< —10dB); therefore, seven NMF components are extracted. To be consistent with NMF features, seven components are extracted in case of FICA and RPCA. The numbers displayed in the last row of each feature vector indicate the component which that feature vector represents. There is no classifier involved in associating each feature to its representing component as the features in Tables 2-4 are clearly associated with the energy density highlights in Fig. 5. For example, the first feature (F_1) in Table 3 is associated with the third component in Fig. 5 (The components in Fig. 5 are numbered 1, 2 and 3 from left to right).

Furthermore, in order to provide a comprehensible demonstration of the extracted feature vectors, the features shown in Tables 2, 3 and 4 are illustrated in Figs. 6(a), 6(b) and 6(c) respectively. Each feature vector is associated with a rectangular region centered at $\bar{n}$ and $\bar{m}$ and width of $\hat{n}$ and $\hat{m}$ in time and frequency respectively, and energy of $\hat{e}_j$.

TABLE 3 7 TF FEATURES EXTRACTED IN EXPERIMENT 2. RPCA IS USED AS THE MD TOOL

Parameters	Features						
	F1	F2	F3	F4	F5	F6	F7
$\hat{e}$	0	-0.5543	-2.9342	-3.6079	-17.2361	-17.6982	-19.5239
$\bar{m}$	0.1230	0.3209	0.1206	0.1522	0.1624	0.2805	0.3143
$\bar{n}$	0.1943	0.1025	0.1974	0.1894	0.1897	0.0893	0.0787
$\hat{m}$	0.0286	0.0530	0.0528	0.0335	0.0673	0.1585	0.1582
$\hat{n}$	0.0064	0.0010	0.0111	0.0077	0.0100	0.0628	0.0559
j	3	2	3	3	3	-	2

TABLE 4 7 TF FEATURES EXTRACTED IN EXPERIMENT 2. FICA IS USED AS THE MD TOOL

Parameters	Features						
	F1	F2	F3	F4	F5	F6	F7
$\hat{e}$	0	-3.1425	-3.3637	-6.1000	-8.7705	-8.8756	-43.1515
$\bar{m}$	0.4493	0.1218	0.1443	0.1038	0.2993	0.1700	0.0852
$\bar{n}$	0.0515	0.1929	0.1903	0.1957	0.1024	0.1812	0.1484
$\hat{m}$	0.0146	0.0116	0.0170	0.0194	0.0852	0.0215	0.0359
$\hat{n}$	0.0067	0.0132	0.0118	0.0197	0.0051	0.0190	0.0512
j	1	3	3	3	2	3	-

TABLE 5 7 TF FEATURES EXTRACTED IN EXPERIMENT 2. GRADIENT-BASED NMF IS USED AS THE MD TOOL

Parameters	Features						
	F1	F2	F3	F4	F5	F6	F7
$\hat{e}$	0	-1.5642	-3.8209	-4.2875	-6.4465	-7.4118	-8.0806
$\bar{m}$	0.4488	0.4511	0.1193	0.1397	0.3257	0.1021	0.1678
$\bar{n}$	0.0512	0.0513	0.1964	0.1918	0.1025	0.2006	0.1855
$\hat{m}$	0.0020	0.0049	0.0082	0.0105	0.0450	0.0081	0.0106
$\hat{n}$	0.0025	0.0031	0.0014	0.0019	0.0008	0.0019	0.0021
j	1	1	3	3	2	3	3

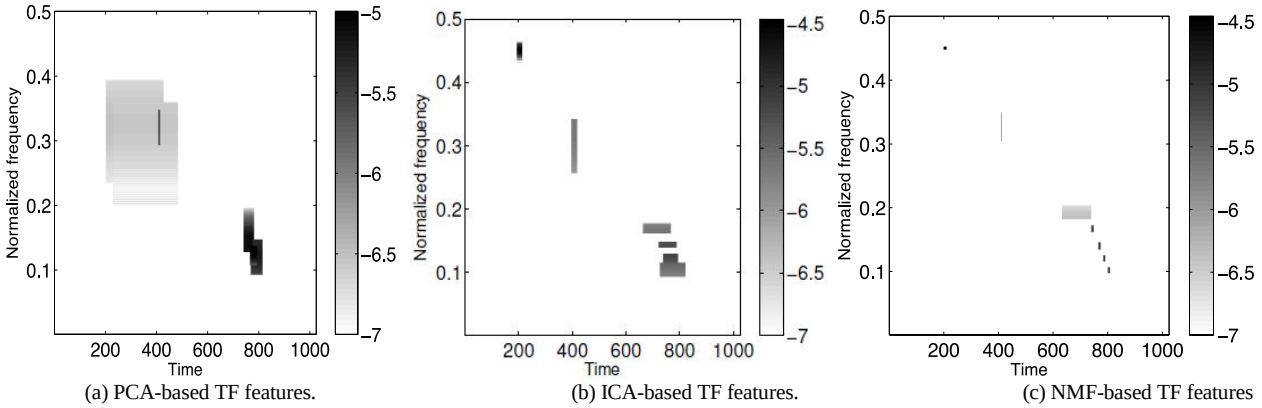
3) Experiment 3: TF Feature Classification

In this experiment, we apply the TF features to a well-known classification problem as introduced in [5, 6]. Test signals are defined as the sum of two linear chirps as defined below:

$$s(t) = \sin\left[2\pi\left(a_0 + a_1 t\right)\right] + \sin\left[2\pi\left(b_0 + b_1 t + b_2 t^2\right)\right], \quad t = 0,\ldots,N-1 \tag{12}$$

where a_0, b_0 belong to a Uniform distribution U(0,1), $a_1 = 0.25$, $b_1 = 0.40$, and $N = 1024$ is the signal length. Two classes are generated by selecting b2 from one of the following Uniform distributions:

$$\text{Class1}: \text{U}\left(\frac{-0.30}{2(N-1)}, \frac{-0.20}{2(N-1)}\right)$$
$$\text{Class2}: \text{U}\left(\frac{-0.15}{2(N-1)}, \frac{-0.05}{2(N-1)}\right)$$



(a) PCA-based TF features. (b) ICA-based TF features. (c) NMF-based TF features

Fig. 6 Result in Experiment 2: The extracted features are plotted in the TF plane. Each rectangle represents one feature vector

A total number of 1100 signals are generated in each class, and TF feature extraction and classification is performed as follows: (i) TF representation (i.e., TF matrix) of each signal is constructed. (ii) A matrix decomposition method is applied to the

TF matrix, and 10 base and coefficient components (i.e., **W** and **H**, and r = 10) are computed for each signal. (iii) A feature vector is extracted from each component pair as explained in Eq. 11; i.e., there are 10 feature vectors for each signal and each feature vector contains 5 feature values. (iv) Linear discriminant analysis (LDA) classifier is used to train and then classify the signals in each class. The classifier is trained using 90% samples and classified over all the signals. The probability of a signal belonging to a class is calculated according to the distribution of its representing feature vectors in different classes. Each signal is labeled to belong to the class associated with the maximum representing class. For example, if 8 feature vectors are classified as Class 2 and the remaining 2 feature vectors are grouped as Class 1, the representing signal will be labeled according to the majority group which in this example, is Class 2. Results obtained from this analysis are listed in Table 6.

TABLE 6 CLASSIFICATION RESULTS USING DIFFERENT TF REPRESENTATIONS AND MATRIX DECOMPOSITION METHODS. MP-TFD: MATCHING PURSUIT TFD. SVD: SINGULAR VALUE DECOMPOSITION. *NOISE IS ADDED TO EACH SIGNAL IN EQN. 12. NOISE IS A RANDOM I.I.D. NOISE AND IS GENERATED BY A ZERO-MEAN GAUSSIAN PROCESS WITH VARIANCE OF 2 (I.E., SNR 0)

Method	Noise-Free		Noise-Added*	
	Spectrogram	MP-TFD	Spectrogram	MP-TFD
PCA	99.9	99.9	79.1	97.7%
ICA	100	100	81.7	99.4%
NMF	99.9	99.9	75.4	98.9%
SVD	99.9	99.9	74.2	98.5%

In the absence of noise, all the four different matrix decomposition methods produce a very accurate classification result. In addition, both spectrogram and MP-TFD are the TF representation result in similar classification accuracies. However, when noise is added to the signals, the accuracy of the Spectrogram-based TF method degrades while MP-TFD one is able to keep the classification accuracy high with only a very small reduction from noise-free scenario. The reason for such behavior can be explained with a close attention to the nature of matching pursuit (MP) algorithm. Performing matching pursuit on a signal, the most coherent structure of the signal is projected and preserved, while the residue of MP represents the random noise (i.e., incoherent structures) present in the signal is discarded [28, 27]. As shown in [28, 27], MP removes white noise from a noisy speech signal and the main time-frequency structures of the original speech signal have been retained while the non-coherent components of noise are discarded as the residue. Hence, MP performs an automatic denoising on the signal. This property of MP allows the technique to be practical in low SNR signals.

4) MD Selection:

From Tables 2-4, Figs. 5 and 7, and the accuracy rate obtained in the synthetic classification problem some interesting properties are concluded as explained in the following:

- Comparing the localization percentages in Fig. 5 of Experiment 1, it can be seen that RPCA, FICA, and NMF perform almost equally well, except for RPCA, which is not successful for decomposition of transient components.
- Another observation is the limitation of PCA in localization of Component 1. The last row of Table 3 and Fig. 6(a) indicates that none of the PCA features represents Component 1.
- From visual inspection of Fig. 7, it is observed that NMF features are highly localized in TF plane, while PCA and ICA-based features are more spread.
- As shown in Tables 2-4, NMF provides the most efficient data reduction compared to ICA and PCA. In this example, we extracted 7 signatures from the TFD; however, the last three features contain only 1.6% of the total feature energy and only the first 4 features obtained from the PCA method are useful features. The same conclusion applies to ICA, as the last feature in Table 4 is too small (-43.1515 dB) to be considered in the TF classification. It can be observed from Fig. 6 that only NMF results in 7 representative TF features.
- Classification accuracy of the three TFM decomposition features is quite high (Experiment 3) with FICA providing the highest accuracy rate (99.36%) and RPCA offering the lowest (97.66%) rate.

Considering the above observations, the three matrix decomposition methods provide similar classification performance, but NMF provides a higher TF localization compared to FICA and RPCA. In addition to the above observation, the advantage of NMF to ICA and PCA can also be analytically explained as described below:

Since PCA and ICA techniques do not guarantee the non-negativity of decomposed matrices instead of directly using them, their squared values ($\tilde{W}$ and $\tilde{H}$, respectively) had to be used in the TF feature extraction. Hence, instead of extracting features from the original matrix (**V**rec ≈ **WH**), PCA and ICA features are extracted from matrix ($\hat{\tilde{V}}_{rec}$) which is constructed by multiplication of two squared matrices as follows:

$$\hat{\tilde{V}}_{rec} = \tilde{W}\tilde{H} \tag{15}$$

The negative elements of the decomposed matrices cause artifacts in the reconstructed TF matrix, and the extracted features might identify energy components at locations where the signal did not originally have any energy value. An example of wrong TF features is the 6th feature vector in Table 3 and the 7th one in Table 4 caused by ICA and PCA techniques, respectively.

Based on the observations in Experiments 1, 2, and 3, and as our goal to extract the TF features that successfully characterize the TF structure of a signal, we suggest NMF as the MD in our developed system, and use it in the rest of this paper. Table 7 summarizes the properties of the three well-known MD techniques as related to the TFD feature extraction.

TABLE 7 DESIRABLE MD PROPERTIES FOR THE TF FEATURE EXTRACTION. THE MORE ARE THE NUMBER OF THE ASTERISK AT EACH PROPERTY INDICATES THAT THE METHOD IS MORE DESIRABLE WITH RESPECT TO THAT SPECIFIC PROPERTY

Property	Localized Reconstruction	TF Feature Localization	Transient Characterization	Efficient Representation	Classification Accuracy
PCA	*	*	*	*	***
ICA	**	**	***	**	***
NMF	***	***	***	***	***

V. INITIALIZATION OF TFM FEATURE EXTRACTION

A successful TFM feature extraction requires a right initialization stage for the NMF optimization technique. One of the common shortcomings of all the NMF optimization approaches is initialization of **W** and **H** matrices. The NMF decomposition algorithms start with random initialization for **W** and **H**, and iteratively modify these two matrices until the cost function is minimized. However, due to the non-convexity of the cost function in both **W** and **H**, depending on the initial matrices, at each optimization, a different local minimum of the cost function may be achieved. As a result, each time we run the algorithm, NMF might result in a different decomposition output. It was theoretically shown in [29] that under some conditions, the NMF decomposition will be unique, but it is also shown that these conditions are not generally satisfied on the case of a real world data. One solution to this shortcoming is appropriate seeding of **W** and **H** matrices. Various efforts have been performed to find alternate seeding approaches in order to influence the NMF convergence to a desired solution [30, 31], SVD is used to seed NMF matrices. To address the initialization problem of the NMF algorithm, we propose a novel seeding method as is explained in the following.

Our motivation in the proposed NMF seeding method is to use our knowledge about the TF structure of a signal to find suitable initialization values for **W** and **H** matrices. In this paper, we suggest matching pursuit TFD (MP-TFD) as a high resolution and cross-term free TF representation [27]. The MP-TFD of a signal ($x(t)$) is constructed as the addition of the Wigner-Ville distribution (WVD) of I Gabor functions as shown below:

$$V(t,f)=\sum_{i=1}^{I}\left|a_{\gamma_i}\right|^2 WVG_{\gamma_i}(t,f) \tag{16}$$

where $\boldsymbol{V}$ is the MP-TFD, WVG_{γ_i} : is the WVD of a Gabor function G_{γ_i}, and a_{γ_i} is the projection of the temporal signal with respect to the Gabor function. A Gabor function is defined as following:

$$G_{\gamma_i}(t)=\frac{1}{\sqrt{s_i}}g\left(\frac{t-p_i}{s_i}\right)\exp\left[j\left(2\left(\pi f_i t+\varphi_i\right)\right)\right] \tag{17}$$

where $g(t)$ is the primary Gaussian function, s_i controls the width of the Gabor function, p_i defines the temporal placement, and f_i and ϕ_i are the frequency and phase of the Gabor function, respectively.

Using the elementary properties of WVD, the WVD of a Gabor function G_{γ_i} in Eq. 17 can be written as follows:

$$WVG_{\gamma_i}(t,f)=WVg\left(\frac{t-p_i}{s_i},s_i\left(f-f_i\right)\right) \tag{18}$$

where **WVg**(t, f) is the WVD of the Gaussian function g(t) in Eq. 17. As shown in Example 4.18 in [28], in case of Gabor atoms, the WVD of a Gaussian atom is a two-dimensional Gaussian which can be written as follows:

$$WVg\left(t,f\right)=\hat{g}^2(f)g(t)^2 \tag{19}$$

where, $\hat{g}(f)$ is the Fourier transform of $g(t)$. In page 28 of [28], it is shown that the Fourier transform of a Gaussian function is also a Gaussian.

From Eq. 19, matrix **WVg** can be shown as follows:

$$WVg = \begin{bmatrix} \hat{g}^2(1) \\ \hat{g}^2(2) \\ \vdots \\ \hat{g}^2(M) \end{bmatrix} \begin{bmatrix} g^2(1) & g^2(2) & \dots & g^2(N) \end{bmatrix} \tag{20}$$

In the above equation, **WVg** is an $M \times N$ matrix, where N and M are the number of samples in $x(t)$ and the frequency resolution, respectively.

We combine Eqs. 18 and 20 in a way that the MP-TFD in Eq. 16 can be written in a matrix format as follows:

$$\mathrm{V}(t,f) = \sum_{i=1}^{I} \left| a_{\gamma i} \right|^2 \begin{bmatrix} \hat{g}^2(s_i(1-f_i)) \\ \hat{g}^2(s_i(2-f_i)) \\ \vdots \\ \hat{g}^2(s_i(M-f_i)) \end{bmatrix} \begin{bmatrix} g^2(\frac{1-p_i}{s_i}) & g^2(\frac{2-p_i}{s_i}) & \dots & g^2(\frac{N-p_i}{s_i}) \end{bmatrix} \tag{21}$$

Comparing the matrix display of MP-TFD (Eq. 22) with the TFM decomposition shown in Eq. 1, we can see that the Gabor-based MP-TFD provides a decomposition of the TFD with decomposition order of I (i.e. r in Eq. 1 is equal to I). In order to decompose all the coherent structures in a signal, the number of MP iterations has to be very large; for example for a 3s audio signal, 1000 iterations are needed. Therefore, the TFM decomposition that is performed through MP-TFD is a very redundant decomposition (i.e. r is very large). According to the nature of matching pursuit algorithm, at every iteration some portion of the signal energy is modeled with an optimal TF resolution in the TF plane, and the first few iteration contains the best-correlated TF functions selected from the Gabor dictionary. Therefore, we assume that the first few atoms in MP-TFD decomposition provide significant information about the TF structure of the signal, and we propose to use the first r Gabor atoms in Eq. 22 to initialize the NMF algorithm as follows:

$$w_i^{Init} = \left| a_{\gamma i} \right| \begin{bmatrix} \hat{g}^2(s_i(1-f_i)) \\ \hat{g}^2(s_i(2-f_i)) \\ \vdots \\ \hat{g}^2(s_i(M-f_i)) \end{bmatrix}, \qquad \text{for } i = 1,\dots,r \tag{22}$$

$$h_i^{Init} = \left| a_{\gamma i} \right| \begin{bmatrix} g^2(\frac{1-p_i}{s_i}) & g^2(\frac{2-p_i}{s_i}) & \dots & g^2(\frac{N-p_i}{s_i}) \end{bmatrix}$$

The proposed seeding method is based on the TF structure of the signal so it is expected to result in a faster convergence. Additionally, unlike the random initialization that results in a different local minimum of the cost function every time we repeat the algorithm, the proposed initialization technique achieves one unique decomposition output. In order to evaluate the proposed seeding method, we used the MP-TFD based and random initialization methods to decompose an 80 ms of a speech signal. Fig. 7 depicts the mean squared error (MSE) between the original and the reconstructed TFM using each initialization methods. As shown in this figure, the proposed MP-TFD based initialization method starts with a larger MSE compared to the random initialization, but it converges much faster than the random seeding. After 8 iterations, NMF with MP-TFD initialization reaches to MSE of 0.01, while the random initialization method requires 18 iterations to achieve the same MSE. This experiment showed that the proposed initialization technique speeds up the convergence of NMF to the desired decomposed matrices.

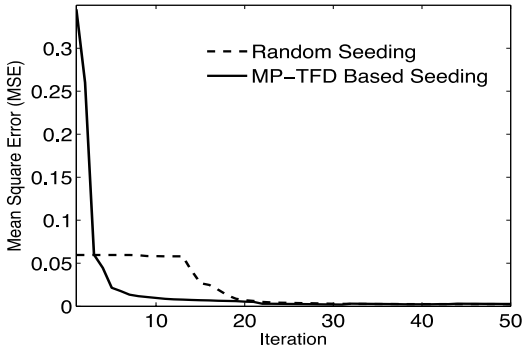


Fig. 7 NMF Convergence is compared for MP and random-based seeding. Using the proposed initialization technique, NMF reaches to a MSE of 0.01 after 8 iterations; while with the random initialization, it takes 18 iterations to achieve the same MSE

VI. APPLICATION OF TFM FEATURE EXTRACTION

The objective of this study is to extract TF features using TFM feature extraction. Such a feature extraction will be helpful in many non-stationary signal feature analysis. In this section, we apply the proposed TFM feature extraction to two real world datasets: pathological voice and audio classification.

A. Pathological Data

Dysphonia or pathological voice refers to speech problems resulting from damage to or malformation of the speech organs. The number of people affected by speech problems is increasing as the modern world places increasing demands on the human voice via mobile telephones, voice recognition software, and interpersonal verbal communications. The purpose of this work is to help patients with pathological problems for monitoring their progress over the course of voice therapy. Currently, patients are required to routinely visit a specialist to follow up their progress. Developing an automated technique saves time for both the patients and the specialist, and can improve the accuracy of the assessments.

We used NMF to perform the proposed TFM decomposition on normal and pathological voices. Once TFM was decomposed into spectral and temporal components, TF features were extracted from each component. K-means clustering was used to locate the pathological features in the feature space. The proposed method was applied to the Massachusetts Eye and Ear Infirmary (MEEI) [32] voice disorders database which consisted of 161 pathological and 51 normal speakers. The sampling frequency of the audio signals was 25 kHz. The TFM decomposition method was applied on 80 ms of continuous speech. r of 15 was chosen to be an appropriate decomposition number for the selected signal length. Four features were extracted from the components with low frequency structures, and five features were derived from the bases with high frequency compositions. More details on the extracted features and the method can be found in our previous work [33]. The trained classifier is tested on all the 212 speeches in the dataset, and a classification accuracy of 98.6% is achieved.

Table 8 demonstrates the accuracy results. From the table, it can be observed that out of 51 normal signals, 50 were classified as normal, and only 1 was misclassified as pathological. Also, the table shows that out of 161 pathological signals, 159 were classified as pathological and only 2 were misclassified as normal. The total classification accuracy of 98.6% is achieved. Some of the high accuracy rates in the literature are listed in Table 9. As can be seen in this table, our proposed TF feature extraction method offers the highest rate reported in the literature using the continuous speech in MEEI database. Some methods (i.e. [34, 35]) report accuracy rates slightly higher than our proposed approach; however, like many other existing methods, their approach is based on sustained vowel (/*a*/), and not continuous speech. Sustained vowels can offer a controlled way of measuring voice characteristics and may produce good results. However, they do not incorporate important vocal function attributes, for example rapid voice onset and termination, fundamental frequency and amplitude variations, and voice breaks. In addition to a high accuracy rate, our algorithm allows the speech pathology recognition be performed on continuous speech which is more challenging but the results are more promising for areal-world speech pathological recognition technique.

TABLE 8 APPLICATION 1: PATHOLOGICAL SPEECH CLASSIFICATION RESULT

Classes	**Normal**	**Abnormal**	**Total**
Normal	50	1	51
Pathological	2	159	161
Normal	98.0%	2.0%	100%
Pathological	1.2%	98.8%	100%

TABLE 9 A SURVEY OF PATHOLOGICAL SPEECH CLASSIFICATION ACCURACY AS REPORTED IN THE LITERATURE

Method	**Year**	**Database (Normal + Pathological)**	**Accuracy Rate**
[36]	2001	Private, sustained vowel (100 + 68)	98.3%
[35]	2002	MEEI, sustained vowel (53 + 657)	98.30%
[37]	2004	MEEI, sustained vowel (53 + 82)	96%
[38]	2005	MEEI, continuous speech (51 + 161)	93.4%
[39]	2005	MEEI, sustained vowel (53 + 77)	95.12%
[40]	2008	Private, words (30 + 30)	98.75%
[41]	2009	MEEI, continuous speech (51 + 161)	97.5%
[34]	2009	MEEI, sustained vowel (53 + 173)	99.69%
[42]	2010	MEEI, sustained vowel (53 + 53)	93.4%

B. Audio Classification

Having approximately 10% of the world population suffering from some sort of hearing loss, one of the important applications of TFM feature extraction is on audio classification in hearing aids for hearing impaired people. Using the proposed feature extraction tool, we are able to discriminate the environmental noise from the desired audio signals, and then prevent the noise signals from being magnified by the hearing aid. We [43] applied the proposed method to a database of 213 audio signals. These

sounds included 23 aircrafts, 17 helicopters, 27 drums, 15 flutes, 31 pianos, 20 animals, 20 birds and 20 insects, and the speech of 20 males and 20 females.

The TF features are derived as follows: First, all the 192 audio signals are transformed into MP-TFD. Next, NMF with decomposition order of 15 (r = 15) decomposes each TFM into 15 base and coefficient vectors. In the present study, experimentally, r = 15 is found to be a suitable choice for our application. Then, 20 features are extracted from each base and coefficient vector. More details on the derived features can be found in [43].

The extracted feature sets for the entire 192 signals were fed to the classifier based on linear discriminant analysis (LDA). Ten-group classification was performed. Table 10 shows the classification accuracy for different classification procedures. In this table, the first column shows the ten classes in the database and the number shows the number of signals in each class; for example, { 'Aircraft' }includes 20 audio signals collected from different aircrafts. The second column in Table 10 shows the classification accuracy with regular LDA. By regular LDA, we mean that 75% of the signal samples in each class are used to train the LDA classifier, and the trained classifier is tested using the entire database.

We compared the accuracy of the TFM decomposition features with the well-known Mel-frequency cepstral coefficient (MFCC) features. MFCCs are short-term spectral features and are widely used in the area of audio and speech processing. In this paper, we compute the first 10 MFCCs for all the segments of the entire length of the audio signals and find the mean and variance of these 10 MFCCs as the MFCC features. For each audio signal we derive 20 features, 10 features are from the mean of the segment MFCCs and the remaining 10 are the variance of the segment MFCCs. These 20 features are computed for all the 192 signals and fed to a LDA-based classifier for classification. The classification results are reported in Table 10. MFCC-based features results in a total accuracy rate of 74.5% which is 11% less than our proposed features.

TABLE 10 APPLICATION 2: ENVIRONMENTAL AUDIO CLASSIFICATION RESULT

Class (#)	TFM Feature Extraction	MFCC
Aircraft (20)	80	80
Helicopter (17)	100	76
Drum (20)	90	60
Flute (15)	100	93
Piano (20)	100	100
Male (20)	90	95
Female (20)	95	90
Animal (20)	55	40
Bird (20)	70	40
Insect (20)	75	70
Total (192)	85.5	74.2

The results from these two real-world applications taken together suggest that the proposed long-term TFM feature extraction is effective in characterizing the non-stationary dynamics of pathological speech and environmental audio signals.

VII. CONCLUSION

This paper presented the time-frequency matrix (TFM) feature extraction based on TFM decomposition and moment-based TF features, which intelligently characterized the long-term information in a nonstationary signal. TFM decomposition is an adaptive approach that was applied to the entire data without any need to segment the data into short durations. Considering the hardware limitations, we had to select a windowed signal; for example, 80 ms for pathological speech classification and 3 s for audio recognition. 80 ms or 3 s as the analysis duration of audio signals are much longer than the duration used in the conventional approaches. Compared to the traditional short-term analysis techniques which have to be performed on 23 to 30 ms of the audio signals, and considering the potential of the proposed framework for application on long signals, the proposed TF feature extraction framework can be relatively categorized as a long-term analysis.

The proposed TF features preserved the time and frequency localization of a given signal, and provided a significant low-dimensional and yet powerful feature extraction tool for real-world signals. The new TF features were unique to the proposed TFM feature extraction framework, which are difficult to be extracted by other methods.

In the present study, we calibrated the proposed TFM feature extraction methodology to further improve the representation of the extracted TF features. Non-negative matrix factorization (NMF) was selected as the desirable matrix decomposition (MD) technique for the proposed TFM feature extraction. The NMF-based TFM feature extraction approach accurately characterized a given TF plane with highly representative and localized TF features. Other contributions of the present paper included integration of matching pursuit-TFD (MP-TFD) algorithm with NMF optimization to seed the decomposed matrices in NMF optimization. This integration improved the time convergence of the algorithm to be half the time required in using the randomly seeding technique. We applied the developed TFM feature extraction to two real-world applications: pathological speech and environmental audio classifications. The high accuracy resulted the proposed technique verified the effectiveness of the proposed TFM feature

extraction methodology compared to the conventional techniques. A significant outcome clearly demonstrated the potential of the new technique as a true non-stationary tool to identify unknown and complex patterns in real-world data.

A FIGURE OF MERIT

The synthetic signal generated using Eq. 8 is composed of seven components, and each component $j(j=1,...,7)$ is characterized with 4 parameters: beginning time, t_{j1}, ending time, t_{j2}, and the line parameters (a_j and b_j) which represent the IF of each component:

$$f_j(t) = a_j t + b_j \tag{23}$$

Fig. 8 demonstrates the IF parameters for a linearly frequency modulated component. For each component, we define the first order moment of the reconstructed TF matrix around its IF as the localization of that component as shown in Eq. 9. As shown in Fig. 8, 2Δ is the frequency interval around the IF in which we calculate the localization. The localization values of each component (j) is calculated using Eq. 9 from the reconstructed TFM (V_{rec}) and the original TFM (**V**): Lcz_j and $Lcz_{O\text{-}j}$, respectively. Finally, the percentage localization of component j is computed as given in Eq. 10. When MD provides an accurate decomposition of the TFM, $| Lcz_j - Lcz_{O\text{-}j} |$ will be a very small value and as a result localizations in Eq. 10 will be very close to 100%, but if MD does not estimate an accurate decomposition, Localizations will be small value.

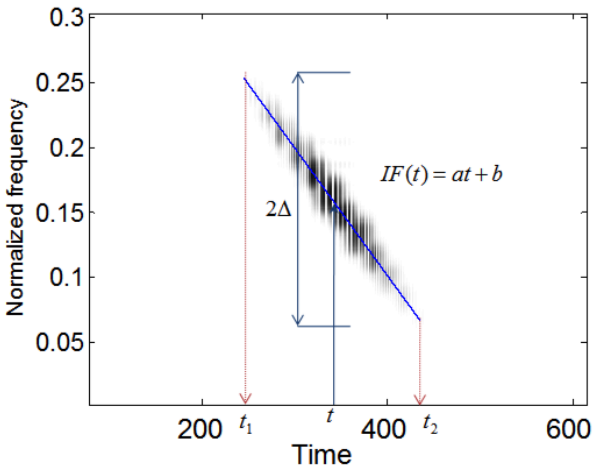


Fig. 8 Each component in the synthetic signal is defined with 4 parameters: start time t1, ending time t2 and the parameters of the IF line a and b. The beginning time of each component is defined as the sample where the signal's amplitude arrives to the 0.1% of the maximum amplitude. The same rule applies for the ending time

REFERENCES

[1] M. Akay. Time frequency and wavelets in biomedical signal processing. *IEEE Press*, ISBN 0780311477, 9780780311473, 1997.

[2] L. Atlas, L. Owsley, J. McLaughlin, and G. Bernard. Automatic feature-finding for time-frequency distributions. *Proc. IEEE-SP Int. Symposium on Time-frequency and Time-scale Analysis*, pp. 333–336, 1996.

[3] B. Tacer and P. Loughlin. Nonstationary signal classification using the joint moments of time-frequency distributions. *Pattern recognition*, 39: 419–424, 1998.

[4] E. J. Zalubas, J. C. O'Neill, W. J. Williams, and A. O. III Hero. Shift and scale invariant detection. *Proc. IEEE ICASSP'97*, 5:3637˜n–3640, 1997.

[5] M. Davy, C. Doncarli, and G. F. Boudreaux-Bartels. Improved optimization of time-frequency based signal classifiers. *IEEE Signal Processing Letters*, 8: 52–57, 2001.

[6] M. Davy, A. Gretton, A. Doucet, and P. Rayner. Optimized support vector machines for nonstationary signal classification. *IEEE Signal Processing Letters*, 9(12): 442–445, 2002.

[7] E. M. Bernat, W. J. Williams, and W. J. Gehring. Decomposing ERP time˜nfrequency energy using pca. *Clinical Neuro-physiology*, 116(6): 1314–1334, Jun. 2005.

[8] A. Delorme, S. Makeig, M. Fabre-Thorpe, and T. Sejnowski. From single-trial EEG to brain area dynamics. *Neurocomput-ing*, 44-46: 1057–1064, 2002.

[9] M. Morup, L. Hansen, J. Parnas, and S. M. Arnfred. Decomposing the time-frequency representation of EEG using nonneg-ative matrix and multi-way factorization. *Technical report, Informatics and Mathematical Modeling, Technical University of Denmark*, 2006.

[10] T. M. Rutkowski, R. Zdunek, and A. Cichocki. Multichannel EEG brain activity pattern analysis in time-frequency domain with nonnegative matrix factorization support. *International Congress Series*, 1301: 266–269, 2007.

[11] B. Boashash. Time frequency signal analysis and processing: a comprehensive reference. *Elsevier Ltd*, 2003.

[12] P. Honeine, C. Richard, and P. Flandrin. Time-frequency learning machines. *IEEE Transactions on Signal Processing*, 55(7): 3930 –

3936, Jun. 2007.

[13] A. Kandaswamy, C. Sathish Kumar, Rm. Pl. Ramanathan, S. Jayaraman, and N. Malmurugan. Neural classification of lung sounds using wavelet coefficients. *Computers in Biology and Medicine*, 34(6): 523–537, 2004.

[14] P. Smaragdis and J. C. Brown. Non-negative matrix factorization for polyphonic music transcription. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 177–180, Oct. 2003.

[15] D. Groutage and D. Bennink. Feature sets for nonstationary signals derived from moments of the singular value decomposition of cohen-posch (positive time-frequency) distributions. *EEE Transactions on Signal Processing*, 48(5): 1498–1503, May 2000.

[16] D. Groutage and D. Bennink. A new matrix decomposition based on optimum transformation of the singular value decomposition basis sets yields principal features of time-frequency distributions. *Proceedings of the Tenth IEEE Workshop on Statistical Signal and Array Processing*, 48: 598–602, Aug. 2000.

[17] J. J. Sylvester. On the reduction of a biline quantic of the nth. order to form of a sum. of n products by a doubles orthogonal substitution. *Messenger of Math*, 19, 42˙n46 1889.

[18] A. Hyvarinen and E. Oja. Independent component analysis: Algorithms and applications. *Neural Networks*, 13, 411-430 2000.

[19] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature 401 (6755)*, 788-791 1999.

[20] M. J. Lee, A. S. Lee, D. Kyungsuk Lee, and Soo-Young Lee. Video representation with dynamic features from multi-frame frame-difference images. *IEEE Workshop on Motion and Video Computing*, 3: 28–35, Feb. 2007.

[21] I. B. Ciocoiu. Occluded face recognition using parts-based representation methods. *Proceedings of the 2005 European Conference on Circuit Theory and Design*, 1:315–318, 28 Aug.–2 Sep. 2005.

[22] H. G. Kim, J. J. Burred, and T. Sikora. How efficient is mpeg-7 for general sound recognition? *In the proceedings of 25th International AES Conference, London, U.K.*, 2004.

[23] N. F. Chikhi, B. Rothenburger, and N. Aussenac-Gilles. A comparison of dimensionality reduction techniques for web structure mining. *IEEE/WIC/ACM International Conference on Web Intelligence*, pp. 116–119, Nov. 2007.

[24] Y. C. Cho, S. Choi, and S. Y. Bong. Non-negative component parts of sound for classification. *In the proceedings of IEEE Int. Symp. Signal Process. Inf. Technol. (ISSPIT), Darmstadt, Germany*, pp. 633–636, 2003.

[25] C. Croux and A. Ruiz-Gazen. Breakdown estimators for principal components: the projection-pursuit approach revisited. *Journal of Multivariate Analysis*, 95, 206-226 2005.

[26] Chih-Jen Lin. Projected gradient methods for nonnegative matrix factorization. *Neural Comput.*, 19(10): 2756–2779, 2007.

[27] S. G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12): 3397–3415, Dec. 1993.

[28] S. Mallat. A wavelet tour of signal processing. *Academic Press*, 1998.

[29] D. Donoho and V. Stodden. When does non-negative matrix factorization give a correct decomposition into parts? *available at http://www-stat.stanford.edu/ donoho*, 2003.

[30] D. Donoho and V. Stodden. Seeding non-negative matrix factorizations with the spherical k-means clustering. *M.S. Thesis, University of Colorado*, 2003.

[31] C. Boutsidis and E. Gallopoulos. Svd based initialization: Ahead start for nonnegative matrix factorization. *Pattern Recogn.*, 41(4): 1350–1362, 2008.

[32] M. Eye E. Infirmary. *Voice disorders database*, (Version 1.03), Lincoln Park, NJ: Kay Elemetrics Corporation.

[33] B. Ghoraani and S. Krishnan. A joint time-frequency and matrix decomposition feature extraction methodology for pathological voice classification. *EURASIP Journal on Advances in Signal Processing*, 2009(ID 928974):11 pages, doi:10.1155/2009/928974, 2009.

[34] P. HenrIquez, J. B. Alonso, M. A. Ferrer, C. M. Travieso, J. I. Godino-Llorente, and F. Diaz de Maria. Characterization of healthy and pathological voice through measures based on nonlinear dynamics. *IEEE Transactions on Audio, Speech, and Language Processing*, 17(6): 1186 – 1195, Aug. 2009.

[35] A. Dibazar, S. Narayanan, and T. W. Berger. Feature analysis for automatic detection of pathological speech. *in: Proceedings of the Second Joint EMBS/BMES Conference, Houston, TX, USA*, 1:182–183, Nov. 2002.

[36] J. Alonso, J. de Leon, I. Alonso, and M. Ferrer. Automatic detection of pathologies in the voice by HOS based parameters. *EURASIP Journal on Applied Signal Processing*, 4: 275–284, 2001.

[37] J. Godino-Llorente and P. Gomez-Vilda. Automatic detection of voice impairments by means of shortterm cepstral parameters and neural network based detectors. *IEEE Transaction on Biomedical Engineering*, 51:380–˙n384, 2004.

[38] K. Umapathy, S. Krishnan, V. Parsa, and D. Jamieson. Discrimination of pathological voices using a time-frequency approach. *IEEE Transaction Biomedical Engineering*, 52:421–˙n430, 2005.

[39] J. Godino-Llorente, P. Gomez-Vilda, N. Saenz-Lech on, M. Blanco-Velasco, F. Cruz-Roldan, and M. A. Ferrer. Discriminative methods for the detection of voice disorders. *in Proceedings of International Conference on Non-Linear Speech Processing*, Apr. 2005.

[40] P. Kukharchik, I. Kheidorov, E. Bovbel, and D. Ladeev. Speech signal processing based on wavelets and svm for vocal tract pathology detection. 5099: 192–199, 2008. 10.1007/978-3-540-69905-7-22.

[41] L. Sugavaneswaran, K. Umapathy, and S. Krishnan. Exploiting the ambiguity domain for non-stationary biomedical signal classification. *In the proceedings of 32nd Annual International IEEE EMBS Conference*, Aug. 31–Sep. 4 2010.

[42] S. Gast˙Un, T. Mar`Ia, and R. Hugo. Pathological voice analysis and classification based on empirical mode decomposition. 5967: 364–381, 2010. 10.1007/978-3-642-12397-9-32.

[43] B. Ghoraani and S. Krishnan. Time-frequency matrix feature extraction and classification of environmental audio signals. *IEEE Transactions on Audio, Speech and Language Processing*, 10.1109/TASL.2011.2118753, 2011.